

The University of Chicago

APPLICATION OF SYMBOLIC METHODS
TO FREQUENCY ARRAYS

A DISSERTATION SUBMITTED TO THE
GRADUATE FACULTY IN CANDIDACY FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF MATHEMATICS

By
HORACE EDWARD WHEELER

Private Edition, Distributed by
THE UNIVERSITY OF CHICAGO LIBRARIES
CHICAGO, ILLINOIS

1931

The University of Chicago

APPLICATION OF SYMBOLIC METHODS
TO FREQUENCY ARRAYS

A DISSERTATION SUBMITTED TO THE
GRADUATE FACULTY IN CANDIDACY FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF MATHEMATICS

By
HORACE EDWARD WHEELER

Private Edition, Distributed by
THE UNIVERSITY OF CHICAGO LIBRARIES
CHICAGO, ILLINOIS

1931



TABLE OF CONTENTS

Chapter	Page
I. INTRODUCTION	
Generating-function concept	1
II. ELEMENTARY THEOREMS	
1. Notation	5
2. Moment theorems	6
3. Linear transformation of scale	7
4. Theorems on combined sampling	7
III. DETERMINATION OF LIMITS FOR INFINITE SEQUENCES OF ARRAYS THROUGH SYMBOLIC METHODS	
5. Lemmas ancillary in the proof of theorem 4	10
6. Arrays on a continuous range - Theorem 4	18
7. Point arrays with continuous-range limit - Theorem 5	23
8. Sequence arrays and limit array denumerable - Theorem 6	27
9. Analogs of the Fourier Integral Theorem	30
10. Continuity property of arrays which arise in combined sampling - Theorem 7	30
11. Foregoing theorems as furnishing sufficient conditions only	31

TABLE OF CONTENTS

Chapter	Page
12. Summary of results	32
IV. APPLICATIONS	
13. Constant continuous probability as a limit of constant point probability	34
14. The point binomial	36
15. Normal probability as a limit in repeated sampling	37
16. Effect of combined sampling where components follow normal law	43
V. FURTHER GENERALIZATIONS	
17. Introduction - The Fourier theorem analogs	45
18. Transformations on the scale of measurement	45
19. Probability of a given sum in sample combination	46
20. The probability of a product of measures by the symbolic method	47



Digitized by the Internet Archive
in 2026

CHAPTER I

INTRODUCTION: THE GENERATING-FUNCTION CONCEPT

This paper considers symbolic representations for probability arrays based on the generating functions of Laplace. As these representations have had a wide though empirical employment in probability, the main body of the discussion is devoted to the writing of sufficient conditions under which symbolic methods apply.

So far as the scheme to be discussed relates to denumerable arrays, it is original with Laplace and is developed by him in volume seven of his Complete Works.¹ An important generalization of the generating-function concept, however, permitting the symbolic representation of continuous arrays as well as denumerable, is due to H. E. Soper and is described by the latter in his monograph "Frequency Arrays."² One other noteworthy contribution, the "method of steepest

¹De Laplace, Théorie Analytique des Probabilités, Oeuvres Complètes, Tome Septieme.

²H. E. Soper, Frequency Arrays; Illustrating the Use of Logical Symbols in Statistical and Other Distributions, Cambridge University Press, 1922.

descents"¹ devised by Fowler for the evaluation of moments and probabilities in statistical mechanics, is to be mentioned in passing, for it is a symbolic method, in important features original, employing the generating functions of Laplace.

The essential idea underlying symbolic representations will become clear if we make a definition of generating functions so generalized that it will apply both to denumerable and to continuous arrays.

Let P_i be a function on a denumerable set (i) to reals. Then $F(T)$ in

$$(1) \quad F(T) \equiv \sum_i T^i P_i$$

is called the generating function of the array

$$P_i \quad (i).$$

Here the function $F(T)$, which represents the array, is defined by a power series the coefficients of which are the functional values in the array. Similarly let $p(x)$ represent a function on a continuous set. Then $f(t)$ in

$$(2) \quad f(t) \equiv \int t^x p(x) dx$$

is said to be the generating function of the array

$$p(x) dx \quad (x, dx).$$

That is, the function $f(t)$ defined by the integral (2) represents the array.

¹R. H. Fowler, Statistical Mechanics, Cambridge University Press, 1929.

The real advantage in symbolic notation is its convenience. Thus in section 2 it will be found that all the moments of an array may be deduced readily from its representation; and in section 4 it will appear that the symbolic notation affords a rapid method for computing the results of repeated or combined sampling.

Obviously unless the arrays $P_i(i)$ and $p(x)dx(x, dx)$ are possessed of certain convergence properties, the generating functions defined by equations (1) and (2) will be non-finite and the representation meaningless, but this question and others of like nature we leave for later consideration.

Soper in his monograph chooses to give to

$$(3) \quad \sum_i T^i P_i$$

a propositional significance. That is, he conceives T^i to represent the proposition that a sample, or event, shall have a measurable property in the amount i . Thus $T^i P_i$ becomes the product of a proposition by the probability of that proposition, and $F(T)$, being the sum of all such products, acquires significance as a description of the sampling possibilities in their entirety. In this interpretation Soper makes no attempt to develop the elaborate postulational basis which would be necessary to render logical the use of symbolic representations in all of the problems to which they apply. Suffice it to say at this point that no systematic analytical consideration of symbolic methods appears in the literature.

The main discourse commences with introductory remarks on notation, next proceeds to the statement of certain well-recognized elementary theorems indicating properties and uses of symbolic expressions and then develops a set of sufficient conditions under which the application of symbolic methods to the more important types of problems may be justified on an analytical basis. The discourse terminates with a summary of results, and several illustrative examples. The final chapter of the paper is devoted to a generalization of the symbolic method.

CHAPTER II

ELEMENTARY THEOREMS

1. Notation. In the introduction, the equation (1),

$$F(T) \equiv \sum_i T^i P_i,$$

was used to define $F(T)$ as generating function of the array P_i (1). We now indicate the same functional relationship from the opposite aspect, namely, " P_i is the coefficient of T^i in $F(T)$," by the notation,

$$(4) \quad P_i \quad T^i: F(T) = T^i: \sum_i T^i P_i \quad (i).$$

Similarly, we express the functional relationship of equation (2), oppositely, by

$$(5) \quad p(x) \equiv t^x: f(t) = t^x: \int t^x p(x) dx \quad (x),$$

to be read " $p(x)$ is the coefficient of t^x in the integrand of an integral which expresses $f(t)$."

We shall have frequent recourse to the substitution of e^{ξ} in place of T and t in relations such as (1), (2), (4) and (5). Thus, for example, equations (2) and (5) can be rewritten as

$$(6) \quad G(\xi) \equiv \int e^{\xi x} p(x) dx$$

and

$$(7) \quad p(x) \equiv e^{\xi x} : G(\xi) = e^{\xi x} : \int e^{\xi x} p(x) dx.$$

We would read the first and last members of equalities (7)

as "p(x) is the integrand coefficient of $e^{\xi x}$ in $\int e^{\xi x} p(x) dx$."

The advantage in this modification of the symbolic method becomes immediately apparent in theorems 1a and 1b below.

For the theorems given here, let p_i , q_i and $p(x)$ $q(x)$ represent measurable probability functions on denumerable and continuous ranges respectively. Also, for theorems 1a and 1b, let

$$F(\xi) \equiv \sum_i e^{\xi i} p_i$$

and

$$f(\xi) \equiv \int e^{\xi x} p(x) dx$$

for all ξ on the range $(-\alpha \leq \xi \leq \alpha > 0)$, define analytic functions of ξ .

2. Moment theorems. It follows then on the basis of these hypotheses that

$$\begin{aligned} (\text{Thm. 1a}) \quad & \left[\begin{array}{c} n^{\text{th}} \text{ moment of the array } p_i \\ \end{array} \right] \quad (i) \\ & = F(0)^{(n)} = \xi^n : n! \sum_i e^{\xi i} p_i \quad (n) \end{aligned}$$

and that

$$\begin{aligned} (\text{Thm. 1b}) \quad & \left[\begin{array}{c} n^{\text{th}} \text{ moment of the array } p(x) dx \\ \end{array} \right] \quad (x, dx) \\ & = f(0)^{(n)} = \xi^n : n! \int e^{\xi x} p(x) dx^1 \quad (n). \end{aligned}$$

¹The first part of the proof for lemma 1 constitutes a proof of the above theorem 1b.

Theorems 1a and 1b naturally constitute the basis of a convenient method for finding all or any moments of an array.

3. Theorem on transformation of origin and scale of the independent variable.

$$(\text{Thm. 2b}) \quad \int e^{\xi a(x-h)} p(x) dx = \int e^{\xi x'} p\left(\frac{x'}{a} + h\right) \frac{dx'}{a}^1$$

The import of the above theorem becomes clear if we notice that the expression

$$(8) \quad \int e^{\xi x} p(x) dx$$

and the corresponding expression in the right member of theorem 2b are symbolic representations for the same array, on scales of measurement for the independent variable which are linear transforms of each other. The effect of substituting $a(x-h)$ in place of x as the coefficient of ξ in (8) is equivalent to that of performing a linear transformation $x' = a(x-h)$ on the scale of measurement for the array.

4. Theorems on combined sampling. In each of the theorems below we assume that the factors in the left member are absolutely convergent.

$$(\text{Thm. 3a}) \quad \sum_i e^{\xi i} p_i \cdot \sum_j e^{\xi j} q_j = \sum_k e^{\xi k} \left(\sum_i p_i q_{k-i} \right)$$

$$(\text{Thm. 3b}) \quad \int e^{\xi x} p(x) dx \cdot \int e^{\xi y} q(y) dy = \int e^{\xi z} \left(\int p(x) q(z-x) dx \right) dz^2$$

¹A more general form of theorem 2b is discussed in section 18.

²A generalizing extension of this theorem is developed in Chapter V.

The third theorem, as is readily recognized, shows that the representation for combined sampling is simply the product of the representations for the individual samples entering into the combination. That is, for example, if p_i is the probability in case A of a sample with a measurement i , and q_j is the probability in case B of a sample with measurement j ; then as is well-known, the probability that the sum of measures for a sample in the one case and the other will be k , is given by the expression

$$(9) \quad \sum_i p_i q_{k-i},$$

and this expression is precisely the function of k whose array is represented by the right member of equation 3a above.

The fundamental theorem of the subject is this theorem three, since through its use a representation for the combination of a great number of samples can be achieved with facility, and the limits approached by proper non-terminating sequences of sample-combination can be readily inferred. Before the theorem can be used for these purposes with any degree of confidence, however, it will be necessary to show (1) that a given representation corresponds to a unique array and (2) that the limit of an sequence of representations is the representation for the limit of the sequence of arrays so represented. As was stated in the introduction, the development of sufficient conditions under which these required relationships may be demonstrated will be the main purpose of this disser-

tation. In the work to follow, convenience of logical development will largely dictate the arrangement of material; however, so that the results may be readily usable, the concluding summary (section 13) will be made in terms of applications. Throughout, the effort will be not so much to employ the weakest hypotheses possible, as to employ those hypotheses which render the applications simplest and most direct.

CHAPTER III

DETERMINATION OF LIMITS FOR INFINITE SEQUENCES OF ARRAYS THROUGH SYMBOLIC METHODS

5. Lemmas ancillary in the proof of theorem 4. Our first theorem in this chapter has to do with the symbolic representation of arrays on a continuous range. In its demonstration we shall require the assistance of the two lemmas given below:

Lemma 1. If a set of probability functions $p_n(x)$ on the ranges ($n = 1, 2, 3, \dots$) and $(-\infty \leq x \leq \infty)$ to reals, and a probability function $p(x)$ on $(-\infty \leq x \leq \infty)$ to reals satisfy the following two relationships,

H_1). There exist positive constants M , α and γ such that

$$|p_n(x) - p(x)| \leq M e^{-(\alpha + \gamma) \cdot x}$$

for positive integer values of n and for all real values of x ,
and

H_2). The sequences concerned satisfy a limit relation,

$$\lim_{n \rightarrow \infty} \int_{-\infty}^{\infty} e^{\xi x} p_n(x) dx = \int_{-\infty}^{\infty} e^{\xi x} p(x) dx$$

for all ξ such that $(-\alpha \leq \xi \leq \alpha > 0)$; then it will be true that these sequences likewise satisfy a limit relation,

$$\lim_{n \rightarrow \infty} \int_{-\infty}^{\infty} e^{i\phi x} p_n(x) dx = \int_{-\infty}^{\infty} e^{i\phi x} p(x) dx$$

for all real finite values of ϕ .

The force of the theorem is that in H_2) the limit relation is assumed to hold for ξ on a restricted real range only, while in the conclusion, the same limit relation is stated to hold for the corresponding variable $i\phi$ on an unrestricted pure imaginary range.

Proof. First we notice that from H_2) of the hypothesis and from the boundedness and closure of the ξ - interval used, we may write that for δ , ξ and ϵ satisfying

$$(\alpha > \delta > 0), \quad [-(\alpha - \delta) \leq \xi \leq (\alpha - \delta)] \quad \text{and} \quad \epsilon > 0,$$

it will be true that there exists some r such that for $n > r$,

$$(10) \quad \left| \int_{-\infty}^{\infty} e^{\xi x} (p_n(x) - p(x)) dx \right| < \epsilon.$$

Combining the results (10) for two different values of ξ , let us say, ξ and $(\xi + \Delta\xi)$, we have

$$(11) \quad \left| \int_{-\infty}^{\infty} e^{\xi x} \frac{e^{\Delta \xi x} - 1}{\Delta \xi} (p_n(x) - p(x)) dx \right| \leq \frac{2\epsilon}{\Delta \xi},$$

where

$$(12) \quad (0 \leq |\xi| + |\Delta \xi| < \alpha - \delta)$$

In the left member of (11) we next expand

$$\frac{e^{\Delta \xi x} - 1}{\Delta \xi}$$

in power series, and at least provisionally distribute the results under two integral signs, thus:

$$(13) \quad \int_{-\infty}^{\infty} e^{\xi x} \frac{e^{\Delta \xi x} - 1}{\Delta \xi} (p_n(x) - p(x)) dx =$$

$$\int_{-\infty}^{\infty} x e^{\xi x} (p_n(x) - p(x)) dx + \Delta \xi \int_{-\infty}^{\infty} e^{\xi x} x^2 \left(\frac{1}{2!} + \frac{\Delta \xi \cdot x}{3!} + \dots \right) (p_n(x) - p(x)) dx.$$

(a)
(b)

From the property H_1) and the equation (12), it is readily inferred that the integral (13a) is well-defined and convergent. Also when we note that

$$\left(\frac{1}{2!} + \frac{\Delta \xi \cdot x}{3!} + \dots + \frac{(\Delta \xi \cdot x)^m}{(m+2)!} + \dots \right) < e^{\Delta \xi \cdot x},$$

it may be seen that the absolute value of the definite integral in expression (13b) is less than some finite bound N for all ξ satisfying (12).

Now combining equations (11) and (13), we have

$$(14) \quad \left| \int_{-\infty}^{\infty} x e^{\xi x} (p_n(x) - p(x)) dx \right| \leq N \cdot |\Delta \xi| + \left| \frac{2\epsilon}{\Delta \xi} \right|$$

By a choice of $\Delta \xi$ sufficiently small, the term $N \cdot |\Delta \xi|$ on the right above becomes as small as we please. Then for any choice of $\Delta \xi \neq 0$, a choice of n sufficiently large makes the term $\left| \epsilon / \Delta \xi \right|$ as small as we please. Hence since the left member of (14) above contains no $\Delta \xi$, we conclude

$$(15) \quad \lim_{n \rightarrow \infty} \int_{-\infty}^{\infty} x \cdot e^{\xi x} (p_n(x) - p(x)) dx = 0$$

for

$$(16) \quad (-\alpha + \delta < \xi < \alpha - \delta),$$

and since δ may be made as small as we desire without affecting the argument, we may rewrite the condition (16) as

$$(17) \quad (-\alpha < \xi < \alpha)$$

We note next that equations (15) and (17) are analogous to the H_2 of the hypothesis, where $x(p_n(x) - p(x))$ in (15) replaces the $(p_n(x) - p(x))$ of H_2 . Also, we observe that if we diminish the ξ -interval we may write in place of H_1 ,

$$(18) \quad \left| x^k (p_n(x) - p(x)) \right| \leq M_k e^{-\left| \left(\alpha + \frac{\gamma}{2} \right) \cdot x \right|} \quad (-\alpha \leq x \leq \alpha),$$

where k is any positive integer, and M_k is a choosable finite constant corresponding to it. Now drawing the conclusion

analogous to (15) for the new hypothesis, we have

$$(19) \quad \lim_{n=0} \int_{-\infty}^{\infty} x^1 e^{\xi x} (p_n(x) - p(x)) dx = 0$$

for the range specified by (17). Continuing the inductive process, we find in general

$$(20) \quad \lim_{n=\infty} \int_{-\infty}^{\infty} x^k e^{\xi x} (p_n(x) - p(x)) dx = 0$$

for k any positive integer, and for the ξ range, as before, specified by (17).

For $\xi = 0$ (20) becomes

$$(21) \quad \lim_{n=\infty} \int_{-\infty}^{\infty} x^k (p(x) - p_n(x)) dx = 0.$$

With the assistance of the results (21) we now proceed to show that

$$(22) \quad \lim_{n=\infty} \int_{-\infty}^{\infty} e^{\zeta x} (p_n(x) - p(x)) dx = 0$$

where ζ is a complex variable bounded by the condition,

$$(23) \quad |\zeta| < \alpha.$$

First we notice from H_1) that the limitand integral of (22) is well-defined. Then we break it up in such manner that a bounding expression may be written for it. Thus:

$$(24) \quad \left| \int_{-\infty}^{\infty} e^{\zeta x} (p_n(x) - p(x)) dx \right| \leq \sum_{i=0}^m \left| \int_{-\infty}^{\infty} \frac{(\zeta x)^i}{i!} (p_n(x) - p(x)) dx \right| + \left| \int_{-B}^B \sum_{i=m+1}^{\infty} \frac{(\zeta x)^i}{i!} (p_n(x) - p(x)) dx \right|$$

(a)
(b)

$$\left| \int_B \sum_{i=m+1}^{\infty} \frac{(\zeta x)^i}{i!} (p_n(x) - p(x)) dx \right| + \left| \int_{-\infty}^{-B} \sum_{i=m+1}^{\infty} \frac{(\zeta x)^i}{i!} (p_n(x) - p(x)) dx \right|,$$

where m is a positive integer and B is a positive number.

Now the right member of (24) is well-defined, since only a finite number of parts is to be considered and each part is less than

$$(25) \quad \int_{-\infty}^{\infty} e^{|\zeta x|} |p_n(x) - p(x)| dx,$$

which by H_1) is known to converge uniformly with respect to n .

In (24) if B be chosen great enough, the two terms (24c) and (24d) will be arbitrarily small, say less than ϵ_c and ϵ_d , for they are the outer ends of an infinitely-extended integral whose integrand is dominated by the integrand of the uniformly convergent integral (25).

Next, B having been chosen, m may be chosen so that (24b) will become less than an arbitrarily-assigned positive number ϵ_b . This result becomes apparent at once when we observe that the integral (24b) is less than

$$(26) \quad \frac{|\zeta B|^{m+1}}{(m+1)!} \int_{-\infty}^{\infty} e^{|\zeta x|} |p_n(x) - p(x)| dx,$$

which becomes arbitrarily small (uniformly with respect to n) if m be chosen large enough.

Finally, B and m fixed, a choice of n sufficiently

great will make the summation (24a) as small as we please, say less than $(m+1)\epsilon_a$. This inference follows out of the fact that each of the $(m+1)$ integrals in (24a) may be made less than ϵ_a separately by a suitable choice of n , say for $n > n_1$. If n be chosen to exceed the finite upper bound of the n_i 's, each of the integrals in (24a) will be less than ϵ_a , and the sum of them will be less than $(m+1)\epsilon_a$, as desired.

Thus we may write

$$(27) \quad \left| \int_{-\infty}^{\infty} e^{\zeta x} (p_n(x) - p(x)) dx \right| \leq (m+1)\epsilon_a + \epsilon_b + \epsilon_c + \epsilon_d.$$

Obviously, by choosing first B , then m , and finally n , we may make the right member of (27) as small as we please, and the conclusion (22), that the limit as to n of the left member of (27) is zero, follows.

Next let us consider using equation (22) in place of the H_2) of the hypothesis, repeating the various steps outlined above. When we come to the analog of equation (20), however, instead of setting $\zeta = 0$ (as we did ξ in the previous argument), let us give to ζ a pure imaginary value,

$$\zeta = i \eta_1.$$

The analog of (21) is then

$$(28) \quad \lim_{n \rightarrow \infty} \int_{-\infty}^{\infty} x^* (e^{i \eta_1 x}) (p_n(x) - p(x)) dx = 0.$$

Since x and η_1 are real,

$$|e^{i\eta_1 x}| = 1,$$

and hence the coefficient $e^{i\eta_1 x}$ may be retained in the subsequent work without its presence affecting the argument. The conclusion comparable to (22) will be that

$$(29) \quad \lim_{n \rightarrow \infty} \int_{-\infty}^{\infty} e^{(i\eta_1 + \zeta)x} (p_n(x) - p(x)) dx = 0$$

for all values of ζ such that $(|\zeta| < \alpha)$.

Obviously by successive repetitions of the process just outlined we can arrive at the conclusion

$$(30) \quad \lim_{n \rightarrow \infty} \int_{-\infty}^{\infty} e^{i\phi x} (p_n(x) - p(x)) dx = 0,$$

where ϕ is any finite, complex number whose pure imaginary part is less in absolute value than α . Thus we have shown more than the conclusion of our lemma requires.

Lemma 2. We need now a statement of Fourier's Integral Theorem, to serve as lemma 2. W. H. Young's¹ expression of this Fourier theorem is paraphrased hereunder.

Given the hypotheses that

H_1). For every finite interval (p, q) which does not contain the point u_0 the expression

$$\int_p^q g(u) du$$

¹W. H. Young, "On Fourier's Repeated Integral," Proc. of the Royal Soc. of Edinburgh, 1910-11, Vol. 31, p. 560.

exists as a Lebesgue integral,

H_2). On some neighborhood (u_0-p, u_0+p) of the point u_0 , $g(u)$ has bounded variation,

H_3). For some $q > u_0$ and $-r < u_0$, the two integrals

$$\int_q^\infty |g(u)| du \quad \text{and} \quad \int_{-\infty}^{-r} |g(u)| du$$

are finite and well-defined; then it follows that

$$\int_0^\infty dv \int_{-\infty}^\infty g(u) \cos[(u-u_0)v] du = \frac{\pi}{2} [g(u_0+0) + g(u_0-0)],$$

where $g(u_0+0)$ and $g(u_0-0)$ designate the values of the function g on either side of a possible discontinuity at u_0 .

6. Conditions applicable to a sequence of probability functions arrayed on a continuous range. With the assistance of the foregoing lemmas we are now prepared to demonstrate a set of conditions under which the limit of a sequence of functions arrayed on a continuous range may be found from the limit of a corresponding sequence of symbolic representations.

Theorem 4. For a sequence of probability functions $p_n(x)$, and a probability function $p(x)$ satisfying the entire hypothesis of lemma 1,¹ and in addition the following as well:

¹In place of having the hypotheses H_1) and H_2) of lemma 1 given, it will do as well to know that there exists

$$\int_{-\infty}^\infty |p_n(x) - p(x)| dx,$$

and that these functions $p_n(x)$, $p(x)$ defined by lemma 1 satisfy the lemma 1 conclusion.

H_3). The function $f(n, i\phi)$, which is defined by

$$f(n, i\phi) \equiv \int_{-\infty}^{\infty} e^{i\phi x} (p_n(x) - p(x)) dx$$

for all real values of ϕ , yields integrals

$$\int_0^{\infty} |f(n, i\phi)| d\phi \quad \text{and} \quad \int_0^{\infty} |f(n, -i\phi)| d\phi,$$

uniformly convergent for all values of n greater than some finite value n_0 ,

H_4). The expression $(p_n(x) - p(x))$ is of bounded variation in a neighborhood of $x = x_0$, and for each value of n ,

H_5). Any discontinuities in the functions $p_n(x)$ at $x = x_0$ have the property

$$\lim_{n \rightarrow \infty} (p_n(x_0 + 0) - p_n(x_0 - 0)) = 0,$$

H_6 . The function $p(x)$ is continuous at $x = x_0$; it follows that there exists a limit for the sequence $p_n(x_0)$ satisfying

$$\lim_{n \rightarrow \infty} p_n(x_0) = p(x_0)$$

In cases where the properties imposed upon x_0 apply at each x -point, convergence of the sequence of arrays to the symbolically-derived limit-array is affirmed by the theorem.

Proof. Using the complex-variable definition of cosine in connection with the function $f(n, i\phi)$ defined in H_3), we see that

$$\frac{e^{-i\phi x_0} \cdot f(n, i\phi) + e^{i\phi x_0} \cdot f(n, -i\phi)}{2} = \int_{-\infty}^{\infty} \cos \phi (x - x_0) \cdot (p_n(x) - p(x)) dx.$$

It is now readily verified that lemma 1 and H_4) together imply the validity of lemma 2, and we may write

$$(31) (p_n(x_0+0) + p_n(x_0-0) - 2p(x_0)) = \frac{2}{\pi} \int_0^\infty \frac{e^{-i\phi x_0} f(n, i\phi) + e^{i\phi x_0} f(n, -i\phi)}{2} d\phi,$$

as the application of lemma 2 to the problem in hand.

For the next step toward our conclusion, we shall break up the right-hand side of equation (31) and add $(p(x_0+0) - p_0(x_0-0))$ to each member, thus getting for $2(p_n(x_0+0) - p(x_0))$ the bounding expression,

$$(32) \quad 2 \left| (p_n(x_0+0) - p(x_0)) \right| \leq \left| (p_n(x_0+0) - p_n(x_0-0)) \right|_{(a)} \\ + \frac{1}{\pi} \left| \int_0^B [e^{-i\phi x_0} f(n, i\phi) + e^{i\phi x_0} f(n, -i\phi)] d\phi \right|_{(b)} \\ + \frac{1}{\pi} \left| \int_B^\infty [e^{-i\phi x_0} f(n, i\phi) + e^{i\phi x_0} f(n, -i\phi)] d\phi \right|_{(c)}.$$

Here obviously, if a sufficiently large positive value be assigned B , then under H_3), for $n > n_0$, (32c) will be arbitrarily small, say less than ϵ_c . (In (32b) and (32c) since the exponents in the coefficients of the f 's are imaginary, these coefficients are not greater than unity in absolute value.) The choice of B made, the ϕ of the integral (32b) may be thought of as ranging over a bounded, closed interval ($0 \leq \phi \leq B$).

At each point of this interval, according to the conclusion of lemma 1, $|f(n, \pm i\varphi)|$ is less than some arbitrary positive value ϵ_b for any choice of n greater than some n_φ . Hence for a choice of n greater than the certainly-finite upper bound of all n_φ on the interval $(0 \leq \varphi \leq B)$, every $|f(n, \pm i\varphi)|$ on the interval is less than ϵ_b . Finally, we notice that for n sufficiently great, under the assumption H_5 , (32a) will be arbitrarily small, let us say less than ϵ_a .

Thus employing the bounding expressions which we have developed, we may write

$$(33) \quad \left| (p_n(x_0+0) - p(x_0)) \right| \leq \frac{\epsilon_a}{2} + \frac{B}{\pi} \cdot \epsilon_b + \epsilon_c.$$

By a large enough choice of B , the last term of the right member above becomes arbitrarily small. Then, with a sufficiently large value for n , the first two terms of the right member become arbitrarily small. Hence we conclude that

$$(34) \quad \lim_{n \rightarrow \infty} (p_n(x_0+0) - p(x_0)) = 0.$$

Obviously, had we used $p_n(x_0-0)$ in place of $p_n(x_0+0)$, a corresponding conclusion would have been reached, namely, that

$$(35) \quad \lim_{n \rightarrow \infty} (p_n(x_0-0) - p(x_0)) = 0.$$

Since by H_6) $p(x_0)$ is known to be unique, the equations (34) and (35) justify the concluding statement of the theorem.

Cor. to thm. 4. If we assume

$$\int_{-\infty}^{\infty} e^{\xi x} p_1(x) dx = \int_{-\infty}^{\infty} e^{\xi x} p_2(x) dx,$$

where ξ is on the range $(-\alpha \leq \xi \leq \alpha > 0)$, and adjoin the properties of $(p_n(x) - p(x))$ stated in theorem 1 for the difference $(p_1(x) - p_2(x))$, the conclusion follows that at each point x_0 where $p_1(x)$ and $p_2(x)$ are continuous,

$$p_1(x_0) = p_2(x_0).$$

That is, the conditions satisfied, a given representation corresponds to a unique array.

To demonstrate the corollary, we define

$$p_n(x) \equiv p_1(x) \quad \text{and} \quad p(x) \equiv p_2(x).$$

Then it is immediately apparent that the hypotheses of theorem 4 are satisfied, with the possible exception of H_3). The Fourier theorem (lemma 2), however, confirms the convergence of the integrals

$$(36) \quad \int_0^\infty f(n, i\varphi) d\varphi \quad (n)$$

whose properties are considered by H_3), and the fact that $p_n(x)$ is a constant as to n in our case makes all the integrals of the sequence (36) the same, thus demonstrating the required convergence of (36) uniformly with respect to n .

Transcribing from theorem 4, we have the conclusion that a sequence of constants $p_1(x_0)$ approaches $p_2(x_0)$ as a limit. But a sequence of constants approaches that constant as a limit. Hence,

$$p_1(x_0) = p_2(x_0),$$

and the uniqueness of the array corresponding to a given representation is confirmed.

7. Case where the array-sequence is on a denumerable range and limit is on a continuous range. Our next theorem renders the foregoing theory applicable to the case where a sequence of probability arrays on a denumerable range approaches as limit an array on a continuous range.

Theorem 5. The hypotheses are:

H₁). The functions $p(n, x(n, i))$ on $(n = 1, 2, 3, \dots)$ and $(i = \dots, -2, -1, 0, 1, 2, 3, \dots)$ to reals, and the function $p(x)$ on $(-\infty \leq x \leq \infty)$ to reals, satisfy the limit relation,

$$\begin{aligned} \lim_{n \rightarrow \infty} \sum_{i=-\infty}^{\infty} e^{\xi x(n, i)} p(n, x(n, i)) \cdot \Delta_i x(n, i) \\ = \int_{-\infty}^{\infty} e^{\xi x} p(x) dx, \end{aligned}$$

where ξ has the range $(-\alpha \leq \xi \leq \alpha > 0)$ and where $\Delta_i(x)$ is an abbreviation for

$$\Delta_i(x) \equiv x(n, i+1) - x(n, i);$$

H₂). The summation

$$\sum_{i=-\infty}^{\infty} |e^{\xi x(n, i)} p(n, x(n, i)) \Delta_i(x)|$$

is convergent and uniformly bounded with respect to n ;

H₃). The quantity $\Delta_i x(n, i)$ has the property

$$\lim_{n \rightarrow \infty} \Delta_i x(n, i) = 0$$

uniformly in i ;

H_4). The function $x(n, i+\theta)$ on $(n = 1, 2, 3, \dots)$ and $(-\infty \leq i+\theta \leq \infty)$ to reals is continuous and properly monotonically increasing with respect to i , and

$$x(n, \infty) = \infty, \quad x(n, -\infty) = -\infty$$

for every n ;

H_5). A function $p(n, x)$ on $(n = 1, 2, 3, \dots)$ and $(-\infty \leq x \leq \infty)$ to reals has as definition,

$$p(n, x) \equiv p(n, x(n, i+\theta)) \equiv p(n, x(n, i)),^1$$

where θ ranges the interval $(0 \leq \theta < 1)$. The conclusion now conditioned is that

$$\lim_{n \rightarrow \infty} \int_{-\infty}^{\infty} e^{\xi x} p(n, x) dx = \int_{-\infty}^{\infty} e^{\xi x} p(x) dx$$

for ξ on the interval $(-\alpha \leq \xi \leq \alpha > 0)$. The nature of this theorem may be readily discerned if one notes that for a given value of n , the $p(n, x)$ used by the conclusion is a "stair-step" array function defined from the point-array function $p(n, x(n, i))$ of the hypothesis H_1).

Proof. Consider the case where $\xi \geq 0$. As a preliminary to the proof, first notice from the definition H_5) above that

$$(37) \quad \int_{x(n, i)}^{x(n, i+1)} e^{\xi x} p(n, x) dx = p(n, x(n, i)) \int_{x(n, i)}^{x(n, i+1)} e^{\xi x} dx.$$

¹The properties of $x(n, i+\theta)$ assumed in H_4) make possible this definition of $p(n, x)$ for every x on the range $(-\infty \leq x \leq \infty)$.

Also note that since $e^{\xi x}$ is monotonically increasing,

$$\int_{x(n,i)}^{x(n,i+1)} e^{\xi x} dx \geq e^{\xi x(n,i)} \cdot \Delta_i x(n,i) \geq$$

(38)

$$\int_{x(n,i)}^{x(n,i+1)} e^{\xi(x-\Delta_i x(n,i))} dx = e^{-\xi \cdot \Delta_i x(n,i)} \int_{x(n,i)}^{x(n,i+1)} e^{\xi x} dx.$$

It follows from (38) that

$$(39) (e^{\xi \Delta_i x(n,i)} - 1) e^{\xi x(n,i)} \Delta_i x(n,i) \geq \int_{x(n,i)}^{x(n,i+1)} e^{\xi x} dx - e^{\xi x(n,i)} \cdot \Delta_i x(n,i).$$

Next choosing an arbitrary positive constant ϵ , we see from H_3) that

$$(40) \quad \Delta_i x(n,i) < \epsilon$$

for every i , and for every n greater than some finite value n_ϵ . From equations (37) and (39) with the assistance of this result we are able to write

$$\left| \int_{-\infty}^{\infty} e^{\xi x} p(n,x) dx - \sum_{i=-\infty}^{\infty} e^{\xi x(n,i)} p(n, x(n,i)) \Delta_i x(n,i) \right|$$

(41)

$$\leq (e^{\xi \epsilon} - 1) \sum_{i=-\infty}^{\infty} e^{\xi x(n,i)} \left| p(n, x(n,i)) \right| \cdot \Delta_i x(n,i)$$

for all values of n greater than n_ϵ . The expansion of $(e^{\xi \epsilon} - 1)$

yields

$$(42) \quad (e^{\xi \epsilon} - 1) = \epsilon \left(\frac{\xi}{1!} + \frac{\xi^2 \epsilon}{2!} + \dots \right).$$

Obviously this expression approaches zero as ϵ approaches zero, ξ remaining constant. The summation in the right member of the inequality (41) is bounded uniformly with respect to n under the assumption H_2). Hence we may make the right member of (41) as small as we please by a choice of n sufficiently large. Thus with the aid of H_1), we see

$$\begin{aligned} & \lim_{n \rightarrow \infty} \int_{-\infty}^{\infty} e^{\xi x} p(n, x) dx \\ (43) \quad &= \lim_{n \rightarrow \infty} \sum_{i=-\infty}^{\infty} e^{\xi x(n, i)} p(n, x(n, i)) \Delta_i x(n, i) \\ &= \int_{-\infty}^{\infty} e^{\xi x} p(x) dx, \end{aligned}$$

which when $\xi \geq 0$ gives the relation to be demonstrated. For $\xi \leq 0$ it is only necessary to reverse the order of the foregoing inequalities (38), (39) and (41).

Consider now the variation of n and $(1+\theta)$, in such manner that $x(n, 1+\theta)$ remains constant. (This will always be possible due to the properties of $x(n, 1+\theta)$ assumed in H_4)). From the definition H_5) we shall be able to write

$$(44) \quad p(n, x(n, 1+\theta)) = p(n, x) \quad (n, x),$$

and it follows that

$$(45) \quad \lim_{n \rightarrow \infty} p(n, x(n, i+\theta)) = \lim_{n \rightarrow \infty} p(n, x)$$

for every x on the range $(-\infty \leq x \leq \infty)$.

Thus we see that in cases where a sequence of arrays on a denumerable set approaches an array on a continuous set, the foregoing theorem will serve to make the problem over into one in which only integrals are concerned and to which theorem 4 may be applied. More exactly,

Cor. to thm. 5. If all of the hypotheses of theorem 5 are satisfied, and in addition $H_1)$ of lemma 1, and $H_3), H_4), H_5)$ and $H_6)$ of theorem 4; then

$C_1)$. For every positive integer value of n there will be a value of $(i+\theta)$ corresponding, such that

$$x(n, i+\theta) = x_0,$$

$C_2)$. And for each value of x_0 , it will be true that

$$\lim_{n \rightarrow \infty} (p(n, x_0) - p(x_0)) = 0.$$

The corollary makes possible direct confirmation of the conditioned symbolic solutions.

8. Sequence-arrays and limit-array on a denumerable set.

We now come to the situation where both the sequence of arrays and the limit of the sequence are functions on a denumerable set. If we liked, we might at this point analogize the proofs of lemmas 1 and 2 and theorem 4, each of which in its present form applies to arrays on a continuous set. While such a method was found on investigation by the author to be success-

ful, the Fourier integral theorem of lemma 2 being replaced by the interesting analog,

$$p(\lambda, w) = \lim_{\substack{n \rightarrow \infty \\ m \rightarrow \infty}} \sum_{t=1}^m \frac{e^{-iwt}}{n} \sum_{x=1}^m e^{ixt} p(\lambda, x),$$

the procedure is admittedly tedious. Hence a very much more concise method based on a suggestion of Mr. Bartky will be presented herewith:

Theorem 6. If $p(n, i)$ is a function on $(n = 1, 2, 3, \dots)$ and $(i = \dots, -3, -2, -1, 0, 1, 2, 3, \dots)$ to reals, and $p(i)$ is a similar function on $(i = \dots, -3, -2, -1, 0, 1, 2, 3, \dots)$ to reals, and if in addition $p(n, i)$ and $p(i)$ satisfy the two conditions,

H_1). There is a limit relation,

$$\lim_{n \rightarrow \infty} \sum_{i=-\infty}^{\infty} x^i (p(n, i) - p(i)) = 0,$$

H_2). There exist constants α and β in the order $\beta > \alpha > 0$, for which the summations

$$\sum_{i=-\infty}^{-1} \alpha^i |p(n, i) - p(i)| \quad \text{and} \quad \sum_{i=0}^{\infty} \beta^i |p(n, i) - p(i)|$$

are both convergent for each value of n ; then it is true that

$$\lim_{n \rightarrow \infty} (p(n, i) - p(i)) = 0$$

for $(i = \dots, -3, -2, -1, 0, 1, 2, 3, \dots)$.

Proof. First note, from H_2), that for values of x satisfying

$$(46) \quad \beta > |x| > \alpha > 0,$$

the sum

$$(47) \quad \sum_{i=0}^{\infty} x^i (p(n,i) - p(i))$$

is an analytic function of x ; also, that the sum

$$(48) \quad \sum_{i=-\infty}^{-1} x^i (p(n,i) - p(i))$$

is an analytic function of $1/x$, which in turn is an analytic function of x in the region specified. It is easy to see then that the function

$$(49) \quad f(n,x) \equiv \sum_{i=-\infty}^{\infty} x^i (p(n,i) - p(i))$$

is analytic, in the region for x indicated by (46), and it follows also that $f(n,x)$ is expansible in unique Laurent series in this annular interval. Thus, applying Laurent's theorem, we may write

$$(50) \quad p(n,i) - p(i) = \frac{1}{2\pi i} \oint_{\substack{\text{circle} \\ |x|=r}} \frac{f(n,x) dx}{x^{i+1}}$$

for each n , each i and for any chosen value of r such that $(\beta > r > \alpha)$.

Now, since the circumference of the circle $|x| = r$ on the complex plane is a bounded, closed set of points at each of which H_1) applies, it follows that $f(n,x)$ may be made arbitrarily small, say less in absolute value than ϵ , at all points of the circumference $|x| = r$ by a choice of n greater than some finite value n_ϵ . Thus from (50) we may infer

$$(51) \quad |p(n,i) - p(i)| < \left| \frac{1}{2\pi i} \cdot \frac{2\pi \cdot \epsilon \cdot r}{r^{i+1}} \right|,$$

and from (51) conclude that

$$(52) \quad \lim_{n \rightarrow \infty} (p(n,i) - p(i)) = 0$$

for each value of i .

9. Analogs of the Fourier Integral Theorem. In passing, it is instructive to note that (a) the Fourier integral theorem in its exponential form, namely,

$$p(w) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-iwx} \int_{-\infty}^{\infty} e^{ixt} p(t) dt dx;$$

(b) the limit-sum equation mentioned in the foreword to the above theorem, and (c) the Cauchy-Laurent integral formula for coefficients of the Laurent series¹ (equation (50)) bear a striking and obvious analogy to each other; and that the various methods suggested and used in proving the foregoing theorems are essentially the same in principle.

10. Continuity property of the arrays which arise in combined sampling. Since theorem 4 requires a knowledge about the continuity of a succession of sample combinations, we shall demonstrate that continuity is a property of the sequential arrays for a rather common case:

Theorem 7. Given two arrays $p(x)$ and $q(y)$, which in the process of sample combination give a third array $r(z)$, defined by the relation,

$$r(z) \equiv \int p(x) q(z-x) dx \quad 2 :$$

Let $p(x)$ be bounded and integrable on the interval $(x_1 \leq x \leq x_2)$ and zero elsewhere and let $q(y)$ be uniformly continuous except

¹In this last case, the analogy would be closer if we replaced x by $e^{\frac{1}{2}\pi i}$.

²See theorem 3b.

at a finite number of points on the interval $(y_1 \leq y \leq y_2)$ and zero elsewhere: Then $r(z)$ is continuous, is zero outside of some interval $(y_2 - y_1 + x_2 - x_1)$ in length and makes contact at each end of its interval of definition with the line $z = 0$. If the continuity of $q(y)$ be strengthened from uniform to absolute continuity, with the exceptions noted, then it will be true that in addition $r(z)$ is absolutely continuous.

Proof. Using the properties of the hypothesis, we write

$$(53) \quad \begin{aligned} |\lambda(z + \Delta z) - \lambda(z)| &= \left| \int p(x) (q(z - x) - q(z + \Delta z - x)) dx \right| \\ &\leq 2M \left| \sum_{(a)} \epsilon \cdot (x_i - x_i) + n N \Delta z \right|, \end{aligned}$$

where M is the upper bound of the absolute values of $p(x)$, ϵ is the arbitrary number corresponding to Δz for the continuity of $q(y)$, n is the number of discontinuities for $q(y)$ and N is the upper bound for the absolute values of these discontinuities. Obviously, by a proper choice of Δz , the bounding expression (53a) may be made as small as we please, and hence $r(z)$ must be a continuous function of z . The remaining properties follow without difficulty.

11. Foregoing theorems as furnishing sufficient conditions only. As was pointed out previously, the foregoing theorems are far from being a complete set. They do now answer the question as to when a symbolic solution may be made and when one can not. They give no criterion as to whether or not a given generating function has an array corresponding.

Indeed, since they provide sufficient conditions only, they doubtless do not cover many of the problems to which the symbolic method is applicable. However, once a symbolic solution has been made, if the hypotheses of suitable preceding theorems do apply, then these theorems assure us that the results symbolically obtained are correct and proper, and in this fact lies the usefulness of the developed set.

12. Summary. An important problem to which the symbolic method applies is that of finding the limiting array resultant from a continued process of repeated - or combined - sampling. It will be well at this point for us to outline, in the way of summary, those successive steps by which the symbolic solution for this case may be accomplished.

First, for each array must be given a symbolic representation (for which see section 1, page 5). Next the successive steps of the sample-combination must be represented by the products of individual representations (Theorems 3a, 3b; page 7). Usually, also, it is found necessary to make a linear transformation on the distribution scale of the array as part of each step in order to keep the array from becoming infinitely displaced and/or infinitely dispersed (Theorem 2b, (page 7).

Once the sequence of representations is properly set up so that a limit exists, this limit must be ascertained and interpreted back as an array. Then the hypotheses of theorem

4¹ (page 18), the corollary to theorem 5 (page 27), or theorem 6 (page 28) must be verified for the solution so obtained. The conclusion of the theorem used will serve to indicate that the result secured is precisely that which would have been obtained had the sequence of combination-arrays been employed and the limit of this sequence sought directly. The nature of the arrays considered in respective theorems of the main discourse is tabulated below:

<u>Theorems</u>	<u>Range for the arrays in the sequence</u>	<u>Range for limit array</u>	<u>Page</u>
4 and 7	continuous	continuous	18 and 30
5	denumerable	continuous	23
6	denumerable	denumerable	28

These same theorems are described in the table of contents, but there the listing is in the order of development.

¹In many examples theorem 7 is of assistance in the verification of the theorem-4 hypotheses.

CHAPTER IV

APPLICATIONS

13. Constant continuous probability as a limit of constant point probability. We shall now apply the foregoing theory in a set of illustrative examples. The first of these is a trivial one which will serve to introduce the notation.

Problem 1. Given (a) a constant probability defined on a finite set of points: To find the limit approached as the number of points in the set is increased indefinitely, (b) the denumerable range being maintained between $-\frac{1}{2}$ and $+\frac{1}{2}$.

From the hypothesis (a) above,

$$(54) \quad p(i) = \frac{1}{(n+1)} = e^{\xi i}: \sum_{i=0}^n e^{\xi i} \left(\frac{1}{n+1} \right). \quad 1$$

Changing the spread of values of the point distribution from $(n - 0)$ to $(\frac{1}{2} - (-\frac{1}{2}))$ by transformation on the i of $e^{\xi i}$, we write

$$\sum_{i=0}^n e^{\xi \left(\frac{i}{n+1} - \frac{1}{2} \right)} \cdot \frac{1}{(n+1)} \quad 2$$

¹Reference: Chapter II, section 1.

²Explanation in theorem 2a.

$$= \frac{e^{-\xi/2}}{n+1} \sum e^{\xi i/n+1}$$

1

$$= \frac{1 - e^{\xi}}{1 - e^{\xi/n+1}} \cdot \frac{e^{-\xi/2}}{n+1}$$

$$= \frac{e^{\xi/2} - e^{-\xi/2}}{\left[1 - \left(1 + \frac{\xi}{n+1} + \frac{\xi^2}{2(n+1)^2} + \dots\right)(n+1)\right]}$$

$$= \frac{e^{\xi/2} - e^{-\xi/2}}{\left(\xi + \frac{\xi^2}{2(n+1)} + \dots\right)}.$$

$$\lim_{n \rightarrow \infty} \sum_{i=0}^n e^{\xi \left(\frac{i}{n+1} - \frac{1}{2}\right)} \cdot \frac{1}{n+1}$$

$$= \frac{e^{\xi/2} - e^{-\xi/2}}{\xi} = \int_{-\xi/2}^{\xi/2} e^{\xi x} \cdot 1 \cdot dx.$$

Hence the limit array is unity on the interval between $-\frac{1}{2}$ and $+\frac{1}{2}$ and zero elsewhere.²

¹Sum geometrical series.

²Proof by cor. thm. 5.

14. The point binomial. As a second example in symbolic representation, we shall reproduce herewith the derivation of an approximation to hypergeometric probabilities in the form of the point binomial. While neither the result nor the method is new,¹ insertion of the problem here is justified by its pertinence as an illustration of the foregoing theory.

Problem 2. Given a very large lot containing a relatively small number M of defective units: When the fraction f of the entire lot is withdrawn as a sample, show the probability that m units of the sample will be found defective will be approximately

$$P(f, M, m) \equiv C_m^M f^m (1-f)^{M-m}$$

We shall start from the obviously-applicable formula for repeated sampling-without-replacement,

$$(57) \quad P(N, M, n, m) = \frac{\frac{n!}{(n-m)! m!} \cdot \frac{(N-m)!}{(N-n-M+m)! (M-m)!}}{\frac{N!}{(N-M)! M!}}$$

Here n and N designate the sample size and lot size respectively. The f of the problem statement is obviously

$$(58) \quad f \equiv \frac{n}{N}.$$

In terms of generating functions the probability $P(N, M, n, m)$ is otherwise expressed by

¹This derivation follows the method of one used by Dr. Warty in his lectures on the Theory of Sampling, at the University of Chicago, Autumn 1930.

$$(59) \quad P(N, M, n, m) = \frac{x^M y^m : (1+xy)^n (1+x)^{N-n}}{z^M : (1+z)^N}.$$

Since we contemplate allowing N to grow without limit, it will be well to replace x and z above by x/N and z/N . Making this change, we have

$$(60) \quad P(N, M, n, m) = \frac{x^M y^m : (1+\frac{xy}{N})^n (1+\frac{x}{N})^{N-n}}{z^M : (1+\frac{z}{N})^N}$$

The limit for this expression (60) as N approaches infinity, while M and n/N (that is to say, f) remain finite, is

$$(61) \quad \lim P(N, M, n, m) = \frac{x^M y^m : e^{xyf} e^{x(1-f)}}{z^M : e^z}.$$

Now if the exponentials occurring in the right member of (61) be expanded into power series, it then is possible to pick out the coefficients of $x^M y^m$ and of z^M required respectively for the numerator and denominator in (61). The calculation yields

$$(62) \quad \begin{aligned} \lim P(N, M, n, m) &= \frac{f^m / m! \cdot (1-f)^{M-m} / (M-m)!}{1/M!} \\ &= C_m^M f^m (1-f)^{M-m} \end{aligned}$$

In order to confirm the result secured through symbolic methods, we need only invoke theorem 6 for each of the three symbolic representations of equation (60) separately.

15. Normal probability as a limit in repeated sampling. The next problem well illustrates the utility of the symbolic method.

Problem 3. Given a probability array, uniformly of class C' except at a finite set of points on a bounded interval of definition: If this array gives the probabilities for single sampling, show that the results for repeated sampling-with-replacement approach as a limit the normal probability function.

The representation of the array for a single sample will be

$$(63) \quad f(a, h, \xi) = \int_c^d e^{\xi a(x-h)} p(x) dx \quad (-\alpha < \xi < \alpha > 0)$$

subject to

$$(64) \quad \int_c^d p(x) dx = 1$$

and

$$p(x) \geq 0,$$

where $p(x)$ is a function uniformly of class C' except at a finite number of points on the bounded interval (c, d) and the a and h are constants which control the horizontal scale and the horizontal position of the origin for the represented array (Thm. 2b).

We shall now consider choosing h so that $f_{\xi}(0)$ (the first partial derivative of f with respect to ξ at $\xi = 0$) is zero. Computing the derivative $f_{\xi}(\xi)$, we have

$$(65) \quad \begin{aligned} f_{\xi}(\xi) &= a \int_c^d (x-h) e^{\xi a(x-h)} p(x) dx \\ &= a \cdot e^{-\xi a h} \left(\int_c^d x e^{\xi a x} p(x) dx - h \int_c^d e^{\xi a x} p(x) dx \right). \end{aligned}$$

By (64) at $\xi = 0$ the integral in the second term of the right member, equation (65), is unity. Hence some finite choice of

h will make

$$(66) \quad f_{\xi}(0) = 0$$

Noting now that various powers of the function $f(\xi)$ represent the results of repeated sampling (theorem 3b), we shall next employ Taylor's theorem to get an expansion for the quantity $(f(\xi))^k$:

$$(67) \quad (f(\xi))^k = (f(0) + 0 + f''(0) \frac{\xi^2}{2!} + \dots)^k, \quad (k = 1, 2, 3, \dots).$$

But

$$\begin{aligned} f(0) &= \int_c^d p(x) dx = 1 \\ f''(0) &= \int_c^d a^2 (x-h)^2 p(x) dx \\ &\quad - \quad - \quad - \quad - \quad - \quad - \quad , \end{aligned}$$

so that we may rewrite (67) as

$$\begin{aligned} (68) \quad f^k(\xi) &= \left(1 + \frac{a^2 \xi^2}{2!} \int_c^d (x-h)^2 p(x) dx + \right. \\ &\quad \left. \frac{a^3 \xi^3}{3!} \int_c^d (x-h)^3 p(x) dx + \dots \right)^k. \end{aligned}$$

Finally, choosing

$$(69) \quad a = \frac{1}{\sqrt{k}}$$

and defining

$$(70) \quad \sigma^2 \equiv \int_c^d (x-h)^2 p(x) dx, \quad 1$$

we may write

$$(71) \quad (f(\xi))^k = \left(1 + \frac{\sigma^2 \xi^2}{2!} \cdot \frac{1}{k} + \frac{\xi^3}{3! k^{\frac{3}{2}}} \int_c^d (x-h)^3 p(x) dx + \dots \right)^k$$

¹It is to be noted that the properties of $p(x)$ imply a positive value for the right member of (70).

$$= \left[1 + \frac{1}{k} \left\{ \frac{\sigma^2 \xi^2}{2!} + \frac{1}{k^{1/2}} \left(\frac{\xi^3}{6} \int_c^d (x-h)^3 p(x) dx + \dots \right) \right\} \right].$$

Now let us examine a typical term of the parenthesis (71b) above:

$$(72) \left| \frac{\xi^m k^{-(\frac{n-3}{2})}}{n!} \int_c^d (x-h)^n p(x) dx \right| \leq \left| \frac{\xi^m k^{-(\frac{n-3}{2})}}{n!} (|c|+|d|+|h|)^n \cdot (d-c) \max_{c \leq x \leq d} |p(x)| \right|$$

where ($n = 3, 4, \dots$), and ($k = 1, 2, 3, \dots$). We observe first that from its continuity properties, $p(x)$ must possess a finite maximum on the interval (c, d) . Then it is obvious from (72) that the parenthesis (71b) must be uniformly convergent with respect to k and less than some finite bound, say M , for any particular ξ , and for all k greater than some finite k_ξ .

Hence, letting k approach infinity in (71), we find

$$(73) \quad \lim_{k \rightarrow \infty} f^k(\xi) = e^{\frac{\sigma^2 \xi^2}{2}}$$

As may readily be verified, the function on the right of equation (73) is expressible as

$$(74) \quad e^{\frac{\sigma^2 \xi^2}{2}} = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{\xi x} \cdot e^{\frac{-x^2}{2\sigma^2}} dx. \quad 1$$

¹The equation (74) is usually demonstrated from the relation

$$1 = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{\frac{-t^2}{2\sigma^2}} dt.$$

(Footnote concluded on following page)

Granting for the moment that our symbolic solution is valid, we see from equation (74) that the limit approached by the sequence of arrays under consideration is the normal probability array

$$(75) \quad e^{\frac{-x^2}{2\sigma^2}} dx \quad (x, dx).$$

Before we may regard this result as final, however, we must show the symbolic method applicable by demonstrating that the hypotheses of theorem 4 (with footnote alternate in place of H_1) and H_2) are satisfied. This we proceed to do.

First employing theorem 7 we see that the sequential arrays

$$(76) \quad p_k(x) \quad (x) \quad (k = 2, 3, 4, \dots)$$

exist and are absolutely continuous functions on bounded intervals. Since the absolute continuity of any function implies that the function is of limited variation, it is seen that

$$(p_k(x) - e^{\frac{-x^2}{2\sigma^2}})$$

Use of the linear transformation $t = x - \sigma^2 \xi$ on t in the equation above yields

$$\begin{aligned} 1 &= \frac{1}{\sqrt{2\pi} \sigma} \int_{-\infty}^{\infty} e^{\frac{-(x - \sigma^2 \xi)^2}{2\sigma^2}} dx \\ &= \frac{1}{\sqrt{2\pi} \sigma} \int_{-\infty}^{\infty} e^{\frac{-\sigma^2 \xi^2}{2}} (e^{\xi x} \cdot e^{\frac{-x^2}{2\sigma^2}} dx), \end{aligned}$$

and equation (74) follows at once.

is of limited variation in the neighborhood of each x -point and for each value k . This argument verifies the hypothesis H_4) of theorem 4.

Also, from the properties of the arrays (76) with the help of the Fourier-integral theory we readily show that the $f(k, i\phi)$ for all $(-\infty \leq \phi \leq \infty)$ and for $(k = 1, 2, 3, \dots)$ exist and are summable with respect to ϕ . The results (73) and (74) above now indicate that the conclusion of lemma 1, namely, for our case that

$$(77) \quad \lim_{k \rightarrow \infty} \int e^{i\phi x} (p_k(x) - e^{\frac{-x^2}{2\sigma^2}}) dx = 0$$

for all ϕ on the range $(-\infty < \phi < \infty)$, is satisfied.

Finally, it is required by H_3) of theorem 4 that

$$(78) \quad \int e^{i\phi t} (f(i\phi))^k d\phi$$

be uniformly convergent. In order to demonstrate this property for (78), we shall put to use the differentiability of $p(x)$.

(For convenience below let us assume that $p(x)$ is of class C' on the closed interval (c, d) , though obviously the argument to be used would be just the same for the slightly less restrictive properties assumed in the statement of the problem).

Through integration by parts, we have

$$(79) \quad \begin{aligned} f(i\phi) &\equiv \int_c^d e^{i\phi x} p(x) dx \\ &= \frac{e^{i\phi x}}{i\phi} p(x) \Big|_c^d - \int_c^d \frac{e^{i\phi x}}{i\phi} p'(x) dx. \end{aligned}$$

And hence with a properly-chosen positive number M we may write

$$(80) \quad |f(i\phi)| \leq \frac{M}{|\phi|}$$

If now we consider the two sets of integrals

$$\int_1^\infty |f(i\phi)|^k d\phi \quad \text{and} \quad \int_1^\infty |f(-i\phi)|^k d\phi,$$

where ($k = 2, 3, 4, \dots$), we see at once in consequence of (80) that each set must be uniformly convergent with respect to k .

Thus, the hypotheses of theorem 4 (with the footnote alternate in place of H_1) and H_2) are found to be satisfied, and the conclusion follows that

$$\lim_{k \rightarrow \infty} (p_k(x) - e^{\frac{-x^2}{2\sigma^2}}) = 0$$

for every real value of x . That is to say, the results of the symbolic solution are confirmed.

16. Effect of combined sampling where components follow normal law. This example, like problem 2, is included mainly as a concise demonstration of a familiar principle.

Problem 4. Show that the probability for a sum of measurements on samples from two sources which give normal-law arrays will itself be a normal-law array.

As was indicated in equation (74) of problem 3, one normal-law array of arbitrary origin and arbitrary spread may be represented by

$$(81) \quad e^{\frac{\sigma_1^2 a_1^2 \xi^2}{2}} \cdot e^{-a_1 h_1 \xi} = \frac{1}{\sqrt{2\pi}\sigma_1} \int_{-\infty}^{\infty} e^{\xi a_1 (x-h_1)} e^{\frac{-x^2}{2\sigma_1^2}} dx$$

and the other by

$$(82) \quad e^{\frac{\sigma_1^2 a_1^2 \xi^2}{2}} e^{-a_1 h_1 \xi} = \frac{1}{\sqrt{2\pi} \sigma_2} \int_{-\infty}^{\infty} e^{\xi a_2 (x - h_2)} e^{-\frac{x^2}{2\sigma_2^2}} dx.$$

Theorem 3b shows that a representation for the sum of measurements of a sample from the one source and a sample from the other source is given by

$$\begin{aligned} & e^{\frac{\sigma_1^2 a_1^2 \xi^2}{2}} e^{-a_1 h_1 \xi} \cdot e^{\frac{\sigma_2^2 a_2^2 \xi^2}{2}} e^{-a_2 h_2 \xi} \\ (83) \quad & = e^{\frac{(\sigma_1^2 a_1^2 + \sigma_2^2 a_2^2) \xi^2}{2}} e^{-(a_1 h_1 + a_2 h_2) \xi} \\ & = \frac{1}{\sqrt{2\pi} \sigma} \int_{-\infty}^{\infty} e^{\xi a (x - h)} e^{-\frac{x^2}{2\sigma^2}} dx, \end{aligned} \quad (b)$$

where in the last expression

$$(84) \quad ah \equiv a_1 h_1 + a_2 h_2$$

and

$$(85) \quad \sigma^2 a^2 \equiv \sigma_1^2 a_1^2 + \sigma_2^2 a_2^2$$

By reference to equation (74) we see that expression (b) in the chain of equalities (83) is a representation for a normal law distribution. As no limit processes are involved here, the validity of the symbolic solution is indicated at once by theorem 3b.

CHAPTER V

FURTHER GENERALIZATIONS

17. Introduction - The Fourier theorem analogs. Incidental to the study presented in the foregoing pages, several promising generalizations have suggested themselves. The first of these, dealing with the algebraic and semi-algebraic analogs of the Fourier integral theorem, has been included in the discussion (section 9, page 30) just after theorem 6. Generalization of the symbolic method itself to render it useful in a wider field of applications will be discussed here. No attempt at mathematical rigor will be made in this connection, since this part of the work is in the spirit of a suggestion.

18. Transformations on the scale of measurement. If the probability of a measurement between x and $x + dx$ is $p(x)dx$, we write the representation for the array,

$$(86) \quad f(\xi) = \int e^{\xi x} p(x) dx.$$

If now the scale of measurement be changed according to the transformation,

$$(87) \quad x = x(\phi),$$

the probability of a measurement between x and $x + dx$ (which

remains the same as before and now is expressed as the probability of a measurement between $x(\phi)$ and $x(\phi) + x_\phi d\phi$ is written

$$p(x(\phi))x_\phi d\phi.$$

On a ϕ -scale this array is represented by

$$(88) \quad g(\xi) = \int e^{\xi\phi} \cdot p(x(\phi))x_\phi d\phi,$$

where it should be emphasized the probabilities are the same as before, but the scale of measurement has been altered.

It is easy to verify that (88) might have been secured directly from (86) by a substitution of

$$(89) \quad \phi = \phi(x)$$

in place of x in $e^{\xi x}$, so that

$$(90) \quad \int e^{\xi\phi} p(x) dx = \int e^{\xi\phi} p(x(\phi))x_\phi d\phi = g(\xi).$$

We have already considered this method for shifting and changing the scale of measurement in theorem 1b, but there for linear transformations only.

19. Probability of a given sum in sample combination.

From the theorem 2b (page 7) we know that, if $f(\xi)$ represents the array of probabilities that a sample from a given lot will have a measurement between x and $x + dx$, and $h(\xi)$ represents the similar array for another lot; then the product $f(\xi) \cdot h(\xi)$ represents the array of probabilities that the sum of measurements on a sample from the one lot and a sample from the other lot will be between x and $x + dx$.

20. The probability of a product of measures by the symbolic method. Next consider transforming the range so that the logarithm of the measurement becomes the new horizontal scale.¹ Now, if the product of the transformed representations be taken, the result will be a representation of the array of probabilities that the sum of the logs of the two measurements will be between, say, ϕ and $\phi + d\phi$; that is, such that the product of the measurements will lie between $\text{antilog } \phi$ and $\text{antilog } (\phi + d\phi)$.

On the basis of foregoing theory, the computation in connection with getting the probability of a given product of the two measures from separate samplings is formulated below:

Let $p(x)$ and $q(y)$ be probability functions, which are zero for all values of x and y less than or equal to zero.

Define:

$$(91) \quad \phi \equiv \log x$$

$$(92) \quad f^*(\xi) \equiv \int e^{\xi \cdot \log x} p(x) dx$$

$$(93) \quad h^*(\xi) \equiv \int e^{\xi \cdot \log y} q(y) dy.$$

If now we can interpret the product $f^*(\xi) \cdot h^*(\xi)$ as

$$(94) \quad f^*(\xi) \cdot h^*(\xi) = \int e^{\xi \cdot \log z} r(z) dz,$$

¹The positive and negative parts of the original range must be treated separately, and zero values on the original range must be regarded as possible singularities in the use of this transformation.

then

$$\int e^{\xi \log z} r(z) dz$$

will be the representation for the probability of a given log product

$$\int e^{\xi z} r(z) dz$$

will be the representation for the probability of a given product, and $r(z)dz$ will be the probability itself that the product xy of measurements from the separate samplings (corresponding to the arrays

$$p(x) \quad (x) \text{ and } q(y) \quad (y))$$

will be between z and $z + dz$.

VITA

Horace Edward Wheeler was born on December 30, 1898, in Denver, Colorado. His public school education began at Pueblo, Colorado, in 1908 and terminated with his graduation from the Central High School there, June, 1917. The year following this graduation he spent at Colorado College, and the next four at Stanford University, receiving the B.A. in chemistry from Stanford University in 1921, the California secondary credential in 1922 and the M.A. in chemistry in 1924. He matriculated at the University of Chicago in September, 1928.

At the University of Chicago, the author has studied in mathematics and applied mathematics under Doctors Bliss, Bartky, Lane, Graves, Barnard, Logsdon, Hoyt, Macmillan, Everett and Lunn and to each of these he acknowledges his indebtedness. The author is especially grateful to Mr. Bartky for his wise advice and his kind application of time and attention in supervising the preparation of the thesis.

